DOCUMENT RESUME

ED 235 580 EA 016 141

AUTHOR Allison, Derek J.
TITLE Process Evaluation: Some Summarizing and Integrating Notes on the Organizational Implications of This Form of Teacher Evaluation.
PUB DATE Feb 81
NOTE 18p.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Classroom Observation Techniques; *Evaluation Methods; *Formative Evaluation; Intervention; Rating Scales; *Research Problems; *School Organization; *Summative Evaluation; *Teacher Evaluation; Teacher Improvement; Teacher Supervision

ABSTRACT

Two methodologies for teacher-focused process evaluation--rating scales and systematic observation--are discussed and comments are made about their characteristics and effective utilization to improve teachers' performance. Process evaluation, referring specifically to the act of evaluating what teachers do in their classrooms, may be of formative or summative nature. Formative evaluation can be used to monitor what is happening and serves as the basis for making decisions about modifying teacher behavior. Summative evaluation is a quantitative statement that summarizes how well the teacher has been performing. Rating scales require the observer of events to structure, weigh, and relate many perceptions before reaching a conclusion. Rating systems would seem to be better suited to summative forms of evaluation. The second method, systematic observation, deliberately focuses on a sharply delimited set of process factors or dimensions. Both teacher and supervisor have information on a single aspect of classroom process that yields ideal formative data. The two approaches can exist if the organization ensures that the rating and observation methodologies employed are recognized as having separate and independent purposes and steps are taken to protect the integrity of each. (MLF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235580

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Derek J. Allison

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

PROCESS EVALUATION:
SOME SUMMARIZING AND INTEGRATING NOTES ON THE ORGANIZATIONAL IMPLICATIONS OF THIS FORM OF TEACHER EVALUATION

Derek J. Allison

Faculty of Education
The University of Western Ontario
London, Ontario

February 1981

EA 016 141

PROCESS EVALUATION

1. The broad objective of the supervisory process can be understood as being to maintain and attempt to improve organizational performance by working with and through organizational members. Organizationscan be understood as deliberately designed social systems that pursue a set of goals through the application of a particular technology, or "way of doing things". Schools (and/or school systems) can therefore be comprehended as organizations that attempt to educate defined groups of persons (the major goal) by applying the technology of classroom teaching. It would be simplistic, or perhaps merely bureaucratic, to assume that classroom teaching is the sole technology employed by schools in the pursuit of their major goal, but it is the core technology employed by schools. It follows that the supervisory process in schools will be concerned with what happens in classrooms.

2. The technology of classroom teaching can be characterized as a process of planned interaction between a single teacher, a group of learners, a defined body of knowledge and an acknowledged set of social concepts and behaviours. The maintaining or improvement of the operation of this core technology could concentrate on any one or combination of these elements. In practice it would seem that most school supervisory practices concentrate on the teacher and/or the manner in which the teacher interacts with the students. Specific consideration of teacher characteristics and qualities focusses on what Mitzel (1960:1484) has dubbed 'Presage' factors: consideration of teacher-class interaction, which can include diagnosing student needs, planning

classroom activities, and evaluating interaction periods (lesson and units), as well as the actual interaction phase; what Mitzel (1960:1484) refers to as Process factors.

3. Process evaluation, therefore, can be taken as referring specifically to the act of evaluating what teachers do in their classrooms, the rationale for this being that the supervisor requires knowledge about the manner in which the core technology of the organization is being operated so that (s)he may work with or through the teacher and/or other organizational members in order to maintain or improve school performance. "Evaluation" implies judgement. This may be of a formative or summative nature. Formative evaluation of process factors refers specifically to the assemblage of performance information that can be used to monitor what is happening and as the basis for making decisions about modifying the process. To a degree this can be imagined as 'fine-tuning' the organizational technology. Characteristics of formative evaluation are that it is generally interactive, cumulative and that related decisions will normally be made by persons close to the technological core. Typically this means the teacher and supervisor concerned with the possible involvement of others closely associated with the core technology (other teachers in the department or division and possibly staff consultants). The judgement involved in the formative evaluation translates primarily into the interpretation of supervisor perceptions and observational data through the application of a set of values. Dependant on the methodology employed these values may be prestructured into the observation process, may exist solely in the consciousness of the observer, or may be the result of discussion between supervisor and teacher.

Summative evaluation of classroom process refers to the generation of a qualitative statement which purports to summarize how well the teacher is (or has been) performing. Characteristics are that it is normally required and legitimated by internally derived or externally imposed organizational rules and is used as input for decisions made at executive or policy levels in the organization that are removed (and often isolated) from the core technology. Typically these decisions concern the contract status and/or deployment of the teacher concerned. It is a characteristic of summative evaluation in organizations employing professional or semi-professional occupations that these decisions will often involve external organizations.

4. Some Methodologies of Process Evaluation in Schools.

Given the conceptual appreciation above, there would appear to be many alternate techniques for providing both formative and summative evaluations of classroom process. The possible alternatives are reduced through the custom of concentrating heavily on the individual teacher's contribution to the core technology of schools. A further constraint would appear to be the traditional emphasis that school systems have placed on summative evaluation of teachers as a prime instrument of supervision. A theoretical explanation for this may lie along the direction that, given the nature of the core technology in publicly governed schools, it is the individual teacher over whom local school authorities have potentially the greatest control, while at the same time they have traditionally lacked the personnel resources to adequately 'fine tune' this technology. Hence local concerns regarding organizational performance may often hinge on retaining, redeploying or dismissing individual teachers. The problems

inherent in this philosophy become acute in an age of strong teacher organizations, rapid changes in curriculum, diverse and mobile student populations and uncertain social norms and values and are compounded by a failure to build valid knowledge about the teaching process.

Two widespread methodologies for teacher-focussed process evaluation are worthy of note.

4.1 Rating Scales. Technically the term rating scale refers to a type of printed form used by observers to record their judgements about teacher performance. Accepted usage, however, frequently identifies the instrument with the basic methodology of which it forms the central feature. The printed form used by the observer(s) displays a number of descriptive statements, each of which is coupled to some form of forced choice response scale. Each descriptive statement purports to represent a characteristic, behaviour or condition that is associated with desirable classroom operation. Typically these statements are concerned with teacher related process activities rather than presage features, although it is not uncommon for both aspects to be included. Likely process descriptives could be "rapport with students, ... clarity of presentation,... lesson planning". Possible presage descriptives could be "dress, ... voice,..." and so on.

The response scale associated with each statement requires the observer(s) to place a check mark or other symbol so as to indicate the quality of teacher performance in each area identified. A possible arrangement could require the observer to write a number from 1 to 3 alongside each statement where 1=superior, 2=average and 3=poor. More numbers may be used but each number will here be associated with a given conception of quality. A popular alternative is to use a "Likert" scale embodying a number of 'boxes' each of which is associated with a designated performance level.

In addition to the rating section described the forms typically have space for the name of the teacher and observer and their signatures, the date, time, place, and a space for comments. Provision may also be made for a global assessment of the teacher and/or for a recommendation regarding continued employment.

Rating forms of this kind are used by observers to record and standardize their perceptions of individual teachers. These perceptions may be generated as a result of prolonged or limited observation focussing on classroom interaction displayed in one lesson or teacher behaviour in the school over a much longer timespan. The observer is usually a supervisor, or an administrator acting in a supervisory capacity, but could be a peer, a student or class of students or even an "outsider". Many combinations of content and application are possible.

4.2 Comments on Rating Scales and their Use.

.1 The rating scale method is very widely used and many teachers (and supervisors) will likely regard rating scales and teacher evaluation as being virtually synonymous.

.2 Nonetheless there is little consistency between the scales used in different jurisdictions. This implies that different school systems have different conceptions of what "good classroom process" and "good teachers" are like.

.3 Statements on forms of this kind can be taken as representing a specification of the teacher and teaching qualities required by a system and the rating scales themselves represent a declaration of possible and desirable performance levels. In other words rating scales set performance standards.

.4 The only constant in a process evaluation method using rating scales is the printed form itself. All other factors are variables, including the perception and judgement of the observer. Different observers of the same lesson or teacher will likely perceive differently and make different judgements as to quality. Hence rating scale methodologies are notoriously unreliable.

.5 Furthermore the rating scale approach does not capture data from which teacher behaviours or classroom interactions can be reconstructed and re-evaluated. Hence it can be extremely difficult for an observer to justify his rating to others who perceived things differently.

.6 These characteristics are common in most high inference situations in which the observer of events must structure, weigh, and relate many percepts before reaching a conclusion. This process may also require the observer to extrapolate from observed events conclusions that are hard to justify.

.7 Nevertheless need for high inference does allow the observer to consider information and other clues from a wide variety of sources before making a judgement. This can be a highly desirable characteristic if the observer is aware of the inherent limitations and has much wisdom and experience.

.8 At least two potentially independent sets of values are operative in rating methodologies. The form itself is structured in accord with the conception of the desirable held by those who designed it. (cf. 4.1.3 above). In addition the observers are required to filter their observations and tentative conclusions through their own value schema, which may or may not be congruent with the values structured into the form. Hence the observer/

/evaluator may well be forced to operate under conditions of dissonance. The rating form, for example, may place great emphasis on aspects of classroom management while the observer/evaluator could consider this dimension of classroom process to be less important. Nevertheless (s)he is forced to provide teacher ratings in the relevant categories knowing that they will weigh heavily in the final product. This is particularly the case if the form (as some do) requires the individual ratings to be totalled to obtain an aggregate summative rating. Problems of this nature can be particularly acute when the form has been in use without change for an extended period of time, especially if the designers of the form are ro longer with the organization. Furthermore the possibility of the values structured into the rating form being congruent with the teachers' value system may often be remote.

.9 The possibility of dissonance between the parties involved and the values structured into the rating scale can often be minimized by scales which embody broadly derived values. Statements such as "empathy with students" gain legitimacy from the fact that they are widely accepted in the society at large as denoting desirable teacher characteristics whereas "reinforces students appropriately" is a characteristic tied to a specific pedagogic theory, the validity of which is rejected by some educators. Partly for this reason most rating scales embody a number of presage factors, that are characteristics of the teacher, not the actual teaching process. Such factors are legitimated primarily by social not pedagogic values and are also much more amenable to validation as they tend to persist, whereas a given lesson will end never to occur again.

.10 Rating systems would seem to be better suited to summative forms of evaluation. In fact the actual process of rating would seem to be a summative kind of evaluation in itself: the placing of check marks to indicate quality is usually final and intended to be a summary of performance for a given evaluation cycle or time span. However many of the rating forms in use require that a specific recommendation be made regarding the organizational status of the teacher and a copy is often to be filed with an executive decision maker occupying an office distant from the core technology itself. In addition rating scales are often designed to be completed by administrative personnel with line authority, i.e. those with the authority to recommend sanctions or rewards.

.11 A related point is that rating scale methodologies are economical. The forms are standard, can be duplicated in quantity, purport to be global so no additional 'capital' investment is required. Furthermore the completion of the form consumes little time and yields summative data that do not need to be reprocessed or otherwise worked on prior to executive consideration.

.12 The weaknesses of the rating system approach would seem apparent from the previous points. Teachers providing unsatisfactory performance can be identified by this method, but specific data on which an intervention strategy could be based are generally not available. The method can also identify average and good teachers but in doing so may also serve to deny personnel the opportunity of developmental aid they may wish or need. In other words rating scales have limited utility in formative evaluation. Furthermore the potential for conflict between teacher, evaluator and executive decision maker would seem high due to the high level of inference

involved in making judgements and the inherent unreliability of the approach. Finally appeals or grievances over a teacher's dismissal justified by a rating methodology could well place a school system in a difficult position if the relevant external agencies question the validity of the values underlying the rating form, which is, in itself, likely to be different to that used in other school systems.

5. Systematic Observation.

Whereas rating scales typically require high inference judgements by the evaluator and yield indices of "teacher quality", systematic observation methodologies typically yield low inference and purely quantitative data which often embody no qualitative implications. But perhaps the major difference between rating and category systems is that whereas the rating approach attempts to provide global information on the teacher and the observed classroom process, category systems deliberately focus on a sharply delimited set of process factors or dimensions. Indeed, many category systems are single-factor referenced. The factor may be teacher questioning, presentation style, levels of conceptualization, tone of voice, praise of students, or any other single factor that is considered to be a part of the classroom teaching process. The data gathered are usually concerned with typing the factor being observed and the frequency with which each type of event is used during the observation period. For example a systematic observation methodology concerned with teacher questioning could provide for each teacher question to be classified by level in a predetermined and defined hierarchy and by frequency for each level. A refinement, or an entirely different system, could concentrate on the seating location of the students who are questioned. The possibilities are legion.

5.1 Characteristics of Systematic Observation Methodologies

.1 First and foremost it is obvious that a systematic observation method of process evaluation is completely unsuitable for yielding summative evaluation data. Only a single or a few delimited aspects of the classroom dynamic are considered in any given observation. Furthermore a major rationale of the methodology is that once accurate data on a specific classroom process are captured, the teacher may choose to deliberately change his/her behaviour, thus yielding the baseline data invalid for summative purposes.

.2 As suggested above, this approach to process evaluation yields ideal formative data. Both teacher and supervisor have information on a single aspect of classroom process. The meaning and significance of the data can be discussed and evaluated by both parties and a decision made as to whether an adjustment is desirable. If so an attempt to fine tune can then be made and the systematic process of observation reported to gauge the success of this attempt.

.3 The success of a development attempt as described above rests not just on the willingness and motivation of the participants but on the reliability of the systematic observation methodology itself. It is generally held that most well designed and well based systematic observation instruments are highly reliable. Thus different observers of the same classroom will yield highly similar, even identical, data.

.4 The expertise of the observer is often, however, important in the matter of reliability. Frequently the observers require some training and practice before they can reliably apply a given instrument in the 'field'.

.5 As may have been guessed from the above comments systematic observation methodologies have as much of a research- as a supervisory application. In fact most have been (and still are) developed in university and teacher training institutions. This is in sharp contrast to rating systems which are usually sponsored by school superintendents and teacher/administrator committees.

.6 Furthermore the systematic observation methodologies are still predominately used in university sponsored or affiliated situations, typical applications being research into teaching processes and teacher development strategies as well as pre- and in-service teacher training. A major impediment to school and school system use of these methodologies is of course, the resource expense involved. Supervisors require training, usually in a variety of methodologies as each is so highly focussed, and effective application consumes much supervisor and teacher time.

.7 Furthermore the systematic observation approach is probably ill-suited to the needs of school system executives concerned with the overall performance level of the system. The data are so tightly coupled to intricacies embedded in the core technology as to be irrelevant to their summative concerns.

.8 The suitability of systematic observation protocols for formative evaluation encourages their use in cyclical teacher developmental interventions. The most widely recognized such strategy is that of "clinical supervision". Characteristics of this developmental process are that teacher and supervisor discuss the teacher's classroom behaviour and cooperatively select one or more facets for systematic observation.

Appropriate observation instruments are then selected or designed and then applied. The data are cooperatively inspected and evaluated. As a consequence of this cooperative judgement a decision is made as to whether improvement should be attempted and, if so, what strategies may yield success. These may then be applied prior to a second session of systematic observation using the same instruments. Results are again discussed and the intervention cycle may then end or continue with a new or related facet of classroom process coming under scrutiny. Clearly there are a number of arrangements possible: a team of teachers may be cooperatively involved, the period between observations can be of variable or standardized duration, a team of observers, consultants, and/or other supervisors could be involved and so on. The points of interest are that designs of this kind provide a structured process for formative evaluation and development, and that such processes rely on systematic observation methodologies.

.9 Judgements are derived from values. The values employed in evaluative processes incorporating systematic observation techniques are rarely built into the observation schedule itself. On the contrary the device will normally collect purely empirical data which must be interpreted through value sets after the observation has taken place. Hence it would seem possible that several different, but perfectly valid, judgements can be made of the same set of data using different value sets. One aspect of this characteristic is that the supervisor is not forced to defend his judgements and hence his values, but can seek and examine alternate conclusions, especially those reached by the teacher being observed. Not only does this reduce the potential of evaluator/evaluatee conflict, it also provides opportunities for the development of new perceptions in both minds and can foster mutual respect and trust.

6. Alternatives, Hybrids and Extensions.

6.1 For the reasons stated and implied above it would appear that process evaluation methodologies that are based firmly on either the rating or the systematic observation approach will serve the needs of only one type of evaluation, summative or formative, and, they may serve neither well. A school system that relies solely on periodic evaluations of teacher performance through rating scales may capture data suited for in-system summative purposes but will be handicapped in pursuing formative/developmental objectives unless it invests in a special sub-system to meet such objectives. There is always the option of believing that such processes can be safely entrusted to individual principal initiatives, but without a purposeful allocation of system resources to this end the results are likely to be unequal, spotty and poorly coordinated. There is also the serious complication of goal displacement which could well take the form of individual supervisors attempting to use the arena so created to serve essentially personal ends, which could possibly be subversive to the system or entirely inappropriate for the teachers concerned.

Relying completely on systematic observation methodologies carries similar penalties. Not all teachers can be evaluated/developed in a supervisory cycle, say a single year, unless the system incurs an extremely large addition to its supervisory manpower. The penalties of not doing this, that is attempting to share the supervisory resource equally among all teachers, will likely involve severe psychological, self concept and motivational dysfunctions among and between individual staff and teachers. But most damaging is the failure of these methodologies to provide appropriate summative use data. The data that are collected are essentially valueless

and probably incomprehensible to decision makers buffered from the technical core. Any attempt to adjust the methodologies to serve summative ends will surely subvert the supervisory process and engender distrust among those personally involved.

6.2 Can the two approaches co-exist? Probably, if the organization ensures that the rating and observation methodologies employed are recognized as having separate and independent purposes and steps are taken to protect the integrity of each. Thus the rating scale employed could be used solely for summative purposes and should allow for recording the effectiveness of, and teacher satisfaction with, any systematic observation based interventions experienced during the supervisory cycle. The complete failure of a teacher to benefit from several weeks of process referenced activity could be important summative data. Furthermore there should be some integration of the values structured in the rating form and the systematic observation methodologies available. Nothing but dissonance and inequity can be expected when the rating scale emphasizes presage and school related aspects and virtually ignores specific elements of classroom process while the observation methodologies in place concentrate on highly specific elements of the teaching act.

Possible safeguards could include providing systematic observation opportunities for all teachers who wish to cooperate in a given period, say three years and all teachers who request such evaluation. Restriction of such service to only those judged by the summative process as being in some way marginal tightly couples the two methodologies in a discriminatory and unhealthy manner. It is assumed, of course, that any such teachers will request evaluation through the systematic observation mode. Failure to do

so provides useful formative and summative data. Another safeguard would involve the strict partitioning of the two evaluation methodologies. Supervisors working with teachers in the systematic observation mode should be prohibited from evaluating the same teachers with the rating methodology. Even the data generated should be partitioned and their circulation rigidly controlled. Perhaps all systematic observation records should be shredded at the end of each intervention.

A dual system of this kind eliminates some of the independent disadvantages of the two approaches, but some remain. The validity and reliability of the rating scales used should still be regarded with suspicion. Furthermore the cost of a realistic dual system will still be very high when compared to a ratings methodology.

6.3 Are there alternatives to these two major modes? Perhaps only two and each of these embodies some elements of both. The most popular alternative is probably the anecdotal report. Such devices usually require the observer to describe and analyze the lesson and/or teacher involved and to analyze and critically comment on what has been observed. The analysis will of necessity be based firmly on the supervisor's values and his (her) comprehension of the system's declared process values, if such are readily available. By itself such analysis and comment is of little use for summative purposes and lacks the systematization for formative intervention. The summative concern is often dealt with by requiring the supervisor to provide a global evaluation, usually in a ranked format of some kind. Thus at the end of the anecdotal report the evaluator must judge the teacher as being, say, superior, average or poor. This requirement probably forces many evaluators to make their judgement prior to writing out their critical comments so as to ensure that the one supports the other. Hence there is no

guarantee that the summative data generated are based on process factors: the supervisor may merely dislike the teacher. To some degree, therefore, the anecdotal-rating method also suffers from problems of reliability and validity and shares many of the other characteristics of rating scales in general. In all likelihood it could well be regarded as one of the more primitive types of rating scale.

The other major alternative is to shift attention from presage, process factors by incorporating product elements. The emphasis here is on the outcome and consequences of the teacher's involvement in the core technology of the school. Effective evaluation of this domain is highly problematic and possibly much more complex than estimates of process efficacy. As such the option requires separate in-depth consideration.